DOCUMENT RESUME

ED 408 347 TM 026 636

AUTHOR DesJardins, Stephen L.; And Others
TITLE Studying the Determinants of Student Stopout: Identifying "True" from Spurious Time-Varying Effects.
PUB DATE Jun 94
NOTE 22p.; Paper presented at the Annual Meeting of the Association for Institutional Research (34th, New Orleans, LA, May 29-June 1, 1994).
PUB TYPE Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Academic Persistence; Causal Models; *College Students; Dropouts; *Enrollment; High Risk Students; Higher Education; Mathematical Models; *Racial Differences; Reentry Students; *Stopouts; *Time
IDENTIFIERS Time to Degree; University of Minnesota Twin Cities

ABSTRACT

Rather than studying the structural paths through which variables affect student persistence in education, this paper offers a reduced form model that focuses on precollege, demographic, and certain current achievement and financial aid variables. This approach does not specify structural paths, but it does have the advantage of requiring only information available in student records. The empirical model used is the discrete time hazard model because duration data are collected term-by-term until student stopout. Stopout, for this study, is defined as the first occurrence of non-continuous enrollment. The model is illustrated with data for all 3,556 students entering the University of Minnesota Twin Cities campus as New High School students in 1985. These students were observed post hoc for 22 terms (3 terms per year for just over 7 academic years) to study student exits from the institution. The basic model assumes time-constant effects of the independent variables believed to affect stopout. Results are generally consistent with those from other studies of persistence, but show more time profile detail. Race differences were found to vary over time, but these differences might reflect differences in the institution, the students, or inadequate multivariate controls in other studies. The hazard model permits analysts to examine whether factors influencing student dropout or graduation vary by initial year of enrollment. More detailed analysis of subgroup hazards should help administrators meet the needs of at-risk groups. (Contains two figures and five tables.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND
DISSEMINATE THIS MATERIAL
HAS BEEN GRANTED BY
Stephen L. DesJardins
TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)

ED 408 347

Studying the Determinants of Student Stopout: Identifying "True" From Spurious Time-Varying Effects

Stephen L. DesJardins
Senior Scientist
University of Minnesota-Twin Cities

Dennis A. Ahlburg
Professor, Industrial Relations
University of Minnesota-Twin Cities

Brian P. McCall
Assistant Professor, Industrial Relations
University of Minnesota-Twin Cities

Paper Presented at the 34th Annual Forum
of the Association for Institutional Research
May 29 - June 1, 1994
New Orleans, LA

TM026636

BEST COPY AVAILABLE

Section I: Introduction

Many studies have attempted to test the two major competing theories of persistence: the student integration model (Spady, 1970; Tinto, 1975) and the student attrition model (Bean, 1980; 1981; Price, 1977). These studies support the importance placed on the predictive validity of precollege variables in the student integration model and the importance of factors external to the institution in the attrition model. However, results are mixed on the empirical validity of the structural relationships specified by the models (Cabrera et al., 1993). While the student integration and attrition models have been viewed as competing frameworks, Cabrera et al. (1993) have shown that this is not the case. Important components of the models overlap and other aspects of the models are complementary. Cabrera and colleagues offer an integrated model that yields a better understanding of the persistence process. Emphasis is placed on the structural specification of the psychological and sociological processes underlying persistence behavior.

In this paper we take a different approach. Rather than specifying the structural paths through which variables affect persistence, we offer a reduced form model that focuses on precollege, demographic, and certain current achievement and financial aid variables. While not informing us of the way in which these variables affect persistence (that is, not specifying structural paths), this approach has the advantage of requiring only information available in student records (no special surveys are required). Not specifying the structural model may not be a serious drawback since previous studies have shown significant differences in the structural pathways estimated in the integration and attrition models (for example, Pascarella 1986).

The reduced form approach used here allows us to identify groups who may be at higher risk of not persisting (holding constant other characteristics) and so inform decision-makers about where to focus policy interventions. In addition we investigate the impact of financial aid on persistence rather than using financial attitudes as does the student attrition model. The former is a more concrete policy variable. A major contribution of our paper is the application of a hazard model to student persistence. Although it is widely appreciated that the process of persistence (and departure) is longitudinal:

> researchers have in fact done very little to explore the temporal dimensions of that process... Rather than pursue that possibility, past research has implicitly assumed that the process of student departure is essentially invariant over the course of the student's career (Tinto, 1988, p. 438).

Previous studies have either ignored the exact timing of stopout/dropout or have focused on an arbitrary time frame such as fall-to-fall enrollment for sophomore, junior, or senior year. Hazard

models allow us to incorporate the exact timing of stopout/dropout into the estimation of persistence behavior. Such models fully utilize the type of data on persistence called for by Pascarella (1986).

The approach used in this paper allows us to address questions such as when is a student at greatest risk of stopout?, does the profile of risk differ across groups?, do particular programs and policies (specifically, financial aid) affect persistence?, and do the effects of these variables on persistence change over time?

Before attempting to answer the aforementioned questions (in Section V), we provide a short history of student integration and attrition research (Section II). In Section III we detail the empirical methodology used in this paper and in Section IV we provide information about the sample and variables used in our model. In Section VI we will discuss the role that hazard models play in informing decision-making and policy formulation, especially at the study institution. In the conclusion (Section VII), directions for further research will be outlined.

Section II: Brief Overview of the Research

Beginning with the pioneering work of Spady (1970) researchers have been attempting to derive a theoretical model that will help us better understand student attrition. Spady began the search for a conceptual framework by applying aspects of Durkheim's model of suicide to the study of the student dropout process. Spady postulated that a number of factors indirectly affected a student's propensity to drop out of college. These factors (shared group values, grades, normative congruence, and support from friends and/or relatives) were expected to affect a student's level of social integration in the institution that, in turn, led to student satisfaction. Once satisfied, students were expected to become committed to the institution thereby reducing their probability of dropping out.

Vincent Tinto (1975) elaborated on Spady's notions of social and academic integration. Tinto hypothesized that individual and background characteristics and educational achievements interact to influence a student's institutional and goal commitment. Tinto theorized that students who are goal-oriented are likely to get higher grades and achieve higher levels of intellectual development than students lacking a goal orientation. Goal driven students are expected to become academically integrated thereby increasing goal commitment even further. The reinforcing of academic goal commitment is likely to reduce the chances that a student will leave the institution. Social integration takes place when a student becomes committed to the institutions social system through interactions with other students and faculty. Once socially

integrated, students become even more attached to the social structure of the institution and are thereby less likely to leave.

Since Tinto's work other researchers have contributed to the search for a conceptual framework of student integration and/or attrition. Price (1977) applied a worker turnover model to the study of student attrition. Price's model postulates that the determinants of satisfaction and hence turnover are structural variables under the control of an institution. He elaborated on the notion that grades were akin to pay in the work environment and were therefore an important source of reward. Bean (1980, 1981) used Price's model to develop an industrial model of student attrition. Bean's work was the first to use several attitudinal variables to predict intent to leave. Another important contribution to the literature by Bean was the "operationalization of specific elements within the person-role fit and the social integration variables of Rootman's and Spady's model" (Bean, 1981, p.12).

Pascarella (1980, 1982) specifies how a cluster of structural and organizational variables affects student attrition in a framework similar to the one proposed by Tinto. Pascarella stresses the importance of student-faculty contact, especially informal contact. Student background, organizational, and institutional factors are thought to interact to influence informal contact with faculty members and other educational outcome variables. Thus, "educational outcomes are expected to be the direct antecedents of persistence/withdrawal decisions" (Bean, 1981, p.13).

Recently, a number of retention/attrition studies have been published. Stage (1988) applied a path analytic model using logistic regression to the study of student attrition. Bean (1992) has developed a conceptual model known as "student dependency theory". His model competes with the previously mentioned academic/social integration models and uses a path analytic approach to estimate the correlation structure. Cabrera et al. have done a number of studies on persistence, emphasizing the role of finances (1992) and specifying a structural model of student retention (1993).

Despite its applicability to retention and other educational issues, the approach used in our paper has been little used in educational research (for a list of exceptions see Singer and Willett, 1992). Willett and Singer (1991) attribute this to a lack of a coherent methodology that allows researchers to: a) model longitudinal risk profiles as a function of multiple predictors and b) incorporate both censored and uncensored cases simultaneously. We would add two more features: c) control for unobserved heterogeneity across groups and d) incorporating time-varying

effects. Singer and Willett (1991) discuss the application of hazard models to student dropout and graduation and nicely illustrate their advantages over existing approaches.

Section III: Empirical Methodology

The empirical model we used to study student attrition is the discrete time hazard model since the duration data is collected term- by-term until stopout. In this study stopout is defined as the first occurrence of non-continuous enrollment. Let K be a discrete random variable measuring the number of periods until a stopout occurs. It is assumed that stopout is influenced by the vector of, (possibly) time-varying, explanatory variables $\mathbf{z}_k$. Let $P(K=k \mid K > k-1, \mathbf{z}_1, ..., \mathbf{z}_k)$ represents the conditional probability that stopout occurs in period k given that it has not occurred by k-1, where $\mathbf{z}_1$, ..., $\mathbf{z}_k$ represent the values of the explanatory variables until time k.[1] The standard model for this conditional probability is the (discrete-time equivalent of the) proportional hazards model (See Cox, 1972; Prentice and Gloeckler, 1978; Meyer, 1986,1990; and Han and Hausman, 1990):

$$P(K = k \mid K > k-1, z_1, ..., z_k) = 1 - \exp(-\exp(\alpha_k + \beta z_k)) \qquad (1)$$

where β is a vector of coefficients that measure the effects of the explanatory variables and α_k is a time-varying constant term, k=1,2,3,.... [2]

One drawback of this model is that it assumes that *all* the determinants of stopout are accounted for by the explanatory variables $\mathbf{z}_k$. In the social sciences, this assumption is rarely satisfied. Biased estimates, however, can arise when not all determinants are accounted for by $\mathbf{z}_k$ even if such unobserved determinants are uncorrelated with $\mathbf{z}_k$ (see Lancaster, 1979 and Heckman and Singer, 1984). Thus, it is important to account for unobserved heterogeneity.

A second drawback of the model in (1) is that it assumes that the effects of the explanatory variables are constant over time. Again, if this assumption is false, biased estimates may result. In studies of student stopout, it is plausible that some of the explanatory variables have time-varying effects. For example, high school rank percentile or GPA may have a substantial impact

---

[1] One technical point is that the time-varying regressor z_k must be measurable with respect to the information available at time k-1.

[2] An alternative model would be the "logit" model:
$P(K=k \mid K>k-1, z_1, ..., z_k) = \exp(\alpha_k+\beta z_k)/(1+\exp(\alpha_k+\beta z_k))$.

on stopout early in a student's academic career, but may not be as important a factor as time goes by. The model outlined below generalizes the model in (1) by allowing for time-varying effects and controlling for the effects of unobserved heterogeneity.

To account for unobserved heterogeneity it is assumed that stopout is influenced by another random variable θ where θ is unobserved and distributed independently of $\mathbf{z}_k$. Let G denote the cumulative distribution function (c.d.f.) of θ. For identification purposes the mean of θ is fixed at 1.

Let $P(K=k \mid K > k-1, \mathbf{z}_1, ..., \mathbf{z}_k, \theta)$ represent the conditional probability that the stopout occurs in period k given that it has not occurred in the first k-1 periods of enrollment, the values of the time-varying regressor in periods 1 through k, $z_1, ..., z_k$, and the unobserved variable θ. It is assumed that

$$P(K = k \mid K > k-1, z_1, ..., z_k, \theta) = 1 - \exp(-\exp(\alpha_k + \beta_k z_k)\theta) \quad (2)$$

where β_k measures the (possibly time-varying) effect of $\mathbf{z}_k$ in period k and α_k is again a time-varying constant term, k=1,2,3,.... Model (1) is the special case of (2) where $\beta_k = \beta$ for all k and θ equals 1 with probability one.

Model (2) is estimated by both maximum likelihood and non-parametric maximum likelihood estimation (see Heckman and Singer, 1984). McCall [forthcoming] has shown that G is non-parametrically identified so that the latter method is feasible. Since θ is unmeasured it must be "integrated out". Let $S(k \mid z_1, ..., z_k)$ denote the survivor function ($P(K>k \mid \mathbf{z}_1, ..., \mathbf{z}_k)$). Then,

$$S(k \mid z_1, ..., z_k) = \int_0^\infty \exp\left(-\sum_{s=1}^{k} \exp(\alpha_s + \beta_s z_s)\theta\right) dG(\theta). \quad (3)$$

The survivor function (3) enters the likelihood for those individuals who graduated at time k+1. For those who stopout at time k,

$$P(K = k \mid z_1, ..., z_k) = S(k-1 \mid z_1, ..., z_{k-1}) - S(k \mid z_1, ..., z_k) \quad (4)$$

enters the likelihood function.

Section IV: Data

The data sample consists of all 3556 students entering the University of Minnesota-Twin Cities campus as New High School (NHS) students in the fall term of 1985. These individuals were observed post hoc for 22 terms (three terms per year for just over seven academic years) with the event under analysis being whether, and if so when, a student exited the institution for one term (stopped out). Of course student exit is potentially a repeatable event (a student could leave then reenter within the 22 term window) but our analysis will concentrate only on the first occurrence of stopout because of the complexity of dealing with repeated events. Our simplifying definition of stopout may have theoretical merit if one thinks that the reasons for first exit are different from those causing multiple exits.

Table 1

Descriptive Statistics

Variable	Range	Mean	SE
Asian	0 - 1	0.04	0.19
African American	0 - 1	0.02	0.14
Hispanic	0 - 1	0.01	0.1
Female	0 - 1	0.45	0.5
Disabled	0 - 1	0.01	0.08
ACT	11 - 35	23.74	4.51
HS Rank %	1 - 99	69.99	24.2
Metro	0 - 1	0.73	0.44
Enrollment Age	15 - 38	18.36	1.27
Cum GPA *	.98 - 4.00	2.73	6.68
Financial Aid *	0 - 1	0.43	0.49
Athlete *	0 - 1	0.03	0.18

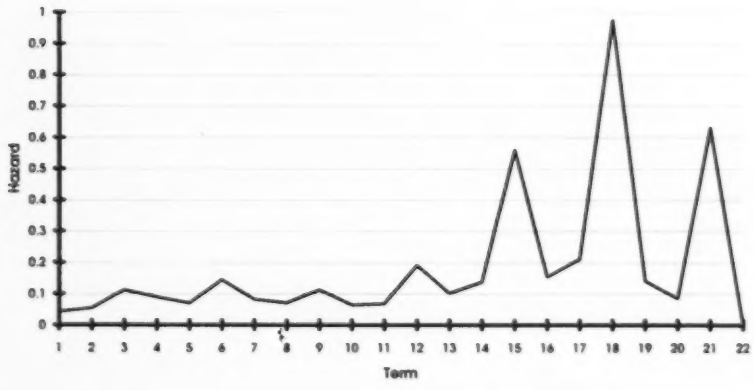
BEST COPY AVAILABLE

As Table 1 indicates, the sample used is composed of 4% Asian Americans, 2% African Americans, and 1% Chicano/Hispanic students. White and International students comprise the ethnicity reference group and account for the other 93% of the sample. About 45% of the sample are females. The average student enters with nearly a 24 on the ACT Composite entrance exam and has graduated in the 70th percentile of their high school class. Most of the entering students (73%) are from the seven county area surrounding Minneapolis/St. Paul. The average student has a first term GPA of 2.73, and about 43% of the sample received financial aid in their first term of enrollment. Only about 3% of the sample are student athletes. As indicated by the *, the last three variables in Table 1 are time-varying regressors thus their values (and means) may change from term-to-term.

Section V: Results

The basic model assuming time-constant effects of the independent variables is shown in Table 2. Column one assumes no unobserved heterogeneity, columns two and three assume unobserved heterogeneity exists and has either a gamma or mass point distribution, respectively. The assumption of no unobserved heterogeneity means that there are no observed or unobserved individual, institutional, or environmental factors that affect stopout other than those specified in the model. Tables 3 through 5 present the results when time varying effects are included and unobserved heterogeneity is controlled for. All Tables show the estimated coefficients and their asymptotic standard errors in parentheses.

Figure 1
Time-Constant Effects Hazard Function



The profile of the baseline hazard of the time-constant effects model assuming no unobserved heterogeneity (Table 2, column 1) is presented in Figure 1. This graph provides a visual

BEST COPY AVAILABLE

description of the times at which students with average characteristics (and time-varying covariates fixed at the mean of term one) are at greatest risk of stopout. It appears that students are most likely to stopout between the spring and fall terms (note the peaks at terms 3, 6, 9, 12...). The wild variation in later periods is probably an artifact of the way the event is defined. Most students miss at least one term therefore the sample size gets small in the outlying terms. A similar hazard profile was found in research on a cohort of 1984 matriculates to the same institution (DesJardins, 1993).

Table 2
Time Constant Effects Models

Variable Coefficient (se)	No Unobserved Heterogeneity (1)	Gamma Heterogeneity (2)	Mass Point Mixing Heterogeneity (3)
Asian	-.25 * (.11)	-.371 * (.149)	-.403 * (.191)
African American	0.172 (.13)	0.238 (.194)	0.311 (.241)
Hispanic	-0.213 (.155)	-0.096 (.256)	0.175 (.341)
Female	.119 ** (.04)	.159 ** (.058)	.286 ** (.074)
Disabled	-0.258 (.289)	-0.149 (.342)	-0.343 (.437)
ACT	.013 * (.006)	.016 * (.008)	.036 ** (.011)
HS Rank %	-.008 ** (.001)	-.011 ** (.001)	-.014 ** (.002)
Metro	-.066 (.045)	-.099 (.064)	-.188 * (.08)
Enrollment Age	.078 ** (.011)	.155 ** (.019)	.218 ** (.026)
Cum GPA	-.009 ** (.0004)	-.012 ** (.0006)	-.017 ** (.001)
Financial Aid	-.158 ** (.041)	-.261 ** (.052)	-.371 ** (.063)
Athlete	-.748 ** (.151)	-1.0 ** (.177)	-1.2 ** (.231)
Likelihood	-8134	-8100	-8066
Likelihood Ratio		68	68
df	-	1	9
Significance	-	> .001	> .001
AIC	16,292	16,226	16,174

The "standard" hazard model (Table 2, column 1) indicates that Whites/Internationals (the reference group), African Americans, and Chicano/Hispanics show no significant difference in their probability of stopout while Asian Americans have a significantly lower probability.

BEST COPY AVAILABLE

Females have a higher probability of stopout than males and students with disabilities do not appear to stopout at rates different from the general population. Two commonly used precollege variables are included: ACT score and high school rank percentile. High school rank percentile has the expected negative effect on stopout while a higher ACT score is associated with a *higher* probability of stopout. It could be that students who have high ACT scores within a given high school rank percentile "test" college for a while and then transfer. Cross-sectional analyses done at the institution tend to support this finding. It seems that these two student ability or "quality" variables are measuring different effects. If this hypothesis is true, then measures of student ability that combine ACT and high school rank percentile for admissions purposes (as happens at the study institution), are inappropriate.

Students from the metro area are commonly believed to be more likely to stopout/dropout than are non-metro students in part because they have better access to metro-area jobs and are thought to be insufficiently integrated into the college community. The results presented here do not support this belief. Students who enter college at a later age appear to be at a higher risk of stopout than those who enter at a younger age. Higher cumulative GPAs and receipt of financial aid tend to lower the probability of stopout. It appears that athletes have a lower probability of stopout than do non-athletes. This finding is contrary to popular belief but as we will show later in the paper, the effect of being an athlete has an important time profile.

Comparing the results in Table 2, we can see that controlling for unobserved heterogeneity matters: the size of the coefficients change, the significance of the metro variable changes, and the value of the likelihood changes. We also calculated Akaike's Information Criterion (AIC), a statistic that adjusts the -2 log likelihood statistic for the number of explanatory variables in the model. Lower values of AIC are indicative of a more desirable model fit. As shown in Table 2, the AIC statistics drop from column 1 through column 3. Thus, we conclude that controlling for unobserved heterogeneity improves the fit of the model. However, when specifying the distribution of the unobserved heterogeneity component, the assumption of gamma heterogeneity may not be appropriate. Imposing a gamma distribution constrains the results, therefore using the more flexible mass point mixing assumption is preferred.

Tables 3 through 5 explore the dynamics of stopout behavior. In these tables two types of significance are established: whether the effect of the variable is significantly different from zero (* = p<=.05, ** = p<=.01) and whether the effect in each year differs from the effect in the first year (coefficients shown in bold). Since the pattern of effects is broadly similar in each table, the discussion will focus on the preferred set of results presented in Table 5. The lower probability

of Asian students stopping out is greatest in year one and then declines over time. African-Americans have a higher probability of withdrawal in the second year but not in any other year. While Ottinger found males more likely to stopout than females (1991) we find that females have higher probabilities of stopout that remains constant over time. Students from a higher high school rank percentile have a constant and lower probability of withdrawal. When other variables are held constant, higher ACT scores lead to a higher probability of stopout in years two and five. The probability of withdrawal for metro-area students is lower in year 3 and years 5+. Older students are at higher risk of stopout particularly in the first four years while students with financial aid are at a constant lower risk of withdrawal than students without aid.

Table 3

Time-Varying Model With No Unobserved Heterogeneity

Variable	Year 1	Year 2	Year 3	Year 4	Year 5 +
Coefficient (se)					
Asian	-0.625 *	-0.036	-0.344	-0.005	**0.32**
	(.27)	(.21)	(.27)	(.26)	(.27)
African American	0.072	0.56 *	-0.71	0.13	0.31
	(.23)	(.24)	(.55)	(.40)	(.50)
Hispanic	0.196	-0.24	0.22	0.13	-1.09
	(.35)	(.48)	(.41)	(.49)	(.72)
Female	0.088	0.18 *	-0.11	0.2	0.24 *
	(.077)	(.08)	(.10)	(.12)	(.12)
Disabled	0.35	-0.38	-1.23	-0.26	-0.46
	(.36)	(.69)	(1.05)	(.76)	(1.15)
ACT	-0.0001	**0.032** **	-0.002	0.017	0.023
	(.01)	(.011)	(.015)	(.017)	(.017)
HS Rank %	-0.007 **	-0.01 **	-0.006 *	-0.008 **	-0.003
	(.002)	(.002)	(.003)	(.003)	(.004)
Metro	-0.053	-0.025	-0.22 *	-0.004	-0.21
	(.086)	(.09)	(.11)	(.13)	(.14)
Enrollment Age	0.11 **	0.09 **	0.061	0.04	**-0.019**
	(.02)	(.023)	(.04)	(.05)	(.05)
Cum GPA	-0.008 **	**-0.014** **	**-0.013** **	-0.006 **	**0.003**
	(.0006)	(.0009)	(.001)	(.001)	(.002)
Financial Aid	-0.028	-0.2 *	**-0.33** **	-0.17	**-0.37** **
	(.08)	(.083)	(.11)	(.12)	(.12)
Athlete	-0.97 **	-0.63 **	-1.67 **	-0.74 *	**.68** **
	(.31)	(.24)	(.53)	(.37)	**(.34)**
Likelihood	-8020				
Likelihood Ratio	228				
df	1				
Significance	>.001				
AIC	16,204				

BEST COPY AVAILABLE

Table 4
Time-Varying Model With Gamma Heterogeneity

Variable	Year 1	Year 2	Year 3	Year 4	Year 5 +
Coefficient (se)					
Asian	-0.628 *	-0.058	-0.47	-0.17	-0.029
	(.29)	(.24)	(.30)	(.30)	(.37)
African American	0.13	0.62 *	-0.71	0.14	0.23
	(.26)	(.31)	(.62)	(.46)	(.68)
Hispanic	0.12	-0.34	0.28	0.088	-1.75
	(.41)	(.57)	(.46)	(.58)	(.92)
Female	0.10	0.25 **	-0.013	0.27 *	0.36 *
	(.08)	(.09)	(.12)	(.13)	(.17)
Disabled	0.3	-0.47	-1.4	-0.64	-1.24
	(.43)	(.80)	(1.12)	(.91)	(1.44)
ACT	0.0004	**0.042** **	0.013	0.03	0.039
	(.01)	(.013)	(.017)	(.02)	(.024)
HS Rank %	-0.008 **	-0.013 **	-0.011 **	-0.016 **	-0.01
	(.002)	(.002)	(.003)	(.004)	(.005)
Metro	-0.07	-0.034	-0.24	-0.038	-0.33
	(.096)	(.10)	(.13)	(.15)	(.18)
Enrollment Age	0.16 **	0.17 **	0.14 **	0.11	0.062
	(.03)	(.031)	(.05)	(.06)	(.073)
Cum GPA	-0.009 **	**-0.017** **	**-0.018** **	-0.012 **	**-0.003**
	(.0007)	(.001)	(.002)	(.002)	(.002)
Financial Aid	-0.12	-0.29 **	**-0.43** **	-0.26	**-0.62** **
	(.08)	(.093)	(.12)	(.13)	(.17)
Athlete	-1.07 **	-0.78 **	-2.05 **	-1.28 **	**.31**
	(.32)	(.27)	(.56)	(.43)	(.44)
Likelihood	-8008				
Likelihood Ratio	184				
df	1				
Significance	>.001				
AIC	16,182				

BEST COPY AVAILABLE

Table 5
Time-Varying Model With Mass Point Mixing Heterogeneity

Variable	Year 1	Year 2	Year 3	Year 4	Year 5 +
Coefficient (se)					
Asian	-1.35 **	-0.73	-0.97 *	-0.54	**-0.015**
	(.45)	(.38)	(.40)	(.37)	(.42)
African American	0.64	0.94 *	-0.47	0.34	0.58
	(.45)	(.42)	(.69)	(.51)	(.84)
Hispanic	0.11	-0.58	0.25	0.038	-1.5
	(.81)	(.71)	(.59)	(.65)	(1.07)
Female	0.23	0.41 **	0.15	0.40 *	0.52 **
	(.15)	(.13)	(.14)	(.15)	(.17)
Disabled	0.80	-0.24	-1.32	-0.78	-0.66
	(.69)	(.99)	(1.16)	(.97)	(1.66)
ACT	0.001	**0.058** **	0.024	0.037	**0.063** *
	(.02)	(.018)	(.019)	(.022)	(.025)
HS Rank %	-0.015 **	-0.021 **	-0.015 **	-0.019 **	-0.013 *
	(.004)	(.003)	(.004)	(.004)	(.005)
Metro	-0.27	-0.17	-0.34 *	-0.068	-0.38 *
	(.17)	(.14)	(.15)	(.17)	(.19)
Enrollment Age	0.24 **	0.24 **	0.19 **	0.16 *	**0.05**
	(.047)	(.031)	(.06)	(.067)	(.063)
Cum GPA	-0.017 **	**-0.027** **	**-0.027** **	-0.02 **	**-0.006** *
	(.001)	(.002)	(.002)	(.002)	(.003)
Financial Aid	-0.41 **	-0.47 **	-0.57 **	-0.36 **	-0.64 **
	(.14)	(.12)	(.14)	(.15)	(.18)
Athlete	-1.69 **	-1.21 **	-2.38 **	-1.71 **	**.16**
	(.55)	(.39)	(.60)	(.52)	(.48)
Likelihood	-7987				
Likelihood Ratio	158				
df	1				
Significance	>.001				
AIC	16,156				

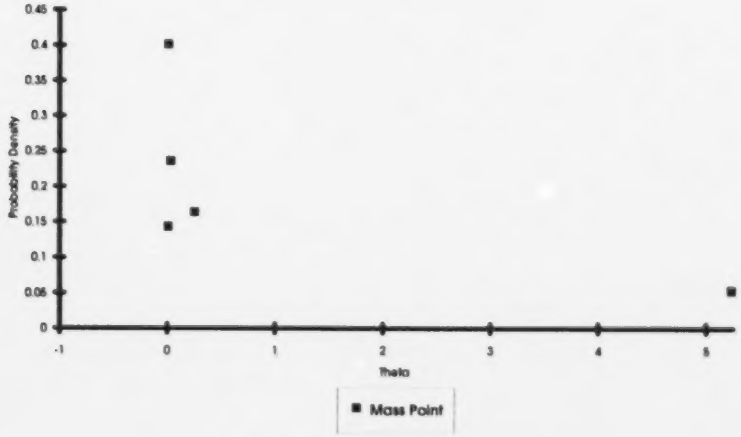
Students who perform better in college have been found to be more likely to persist (Cabrera et al., 1993). In this analysis we also find higher persistence among higher performing students but note that this effect varies over time. The effect of performance seems to be "U" shaped. The probability of withdrawal falls from year one through years two and three and then rises, but remains lower than that of students with lower GPAs. This finding may be important because "grades tend to reflect not only requisite intellectual skills but also desirable personal work habits and attitudes" (Pascarella and Terenzini, 1991, p.388). Because there is so little evidence on the

BEST COPY AVAILABLE

effect of athletic status on educational attainment (Pascarella and Terenzini, 1991), it is worth noting that student athletes are significantly less likely than similar students to stopout during their first four years. However, in the fifth and succeeding years, when many athletes lose their eligibility, they are at a statistically significant higher probability of withdrawing from the institution than in year one.

Regarding the assumptions underlying the unobserved heterogeneity variable in the time-varying models, the gamma distribution does not fit the data as well as the mass point mixing specification. This result can be readily ascertained by comparing the AIC goodness of fit statistics on Tables 4 and 5. The AIC for the gamma heterogeneity model is 16182 compared to 16156 for the mass point mixing model making the latter a more desirable model fit. Also noteworthy is the distribution of unobserved heterogeneity under the mass point mixing model. Figure 2 shows how four of the mass points are grouped near zero with an outlying support point at 5.2. The four mass points near zero reveal that even though most students in the sample are statistically different with respect to unobservable characteristics, there may be no practical significance between them. About 86% of the mass of the unobserved heterogeneity distribution is concentrated in a narrow band (from .0006 to .25) around zero. However, the outlying mass point indicates that some subgroup of students with unobserved characteristics account for 16% of the mass and have a very high probability of stopout.

Figure 2
Heterogeneity Distributions



BEST COPY AVAILABLE

Section VI: Discussion and Policy Implications

Our results are generally consistent with those of other studies of persistence but show more time profile detail. We found race differences as have other researchers, but we found that these differences vary over time. We found that Asians have consistently lower probabilities of withdrawal than other groups, but these differences decline over time. African-Americans are commonly found to have higher probabilities of withdrawal than other groups (Dey & Astin 1989, Ottinger 1991) but we found this to be the case only in the second year of school. We found no differences between Hispanics and Whites/Internationals while Ottinger (1991) concluded that Hispanics are less likely to persist in college. These differences in findings could reflect differences in the retention performance of the study institution, differences in the students at the institution, or inadequate multivariate controls in other studies of persistence.

More detailed analyses of subgroup hazards should help administrators to meet the needs of at risk groups. Research done at the national level indicates that older students, minorities, and individuals from low-income families are prone to higher rates of dropout (American Council of Education, 1991). The recent influx of non-traditional students and the expectation that this trend will continue means that programs will need to be developed to meet their special needs. The results found in this study are consistent with the national level studies in that older students, and students not receiving financial support had higher probabilities of leaving before completion. As some administrators and legislators have suggested, increasing the levels and targeting of financial aid to those in greatest need would help reduce dropout related attrition.

Also, federal government requirements (like the Student-Right-to-Know legislation) to collect and disseminate retention/attrition statistics could benefit from the hazard model approach. The legislation, as written, has come under fire because of its cross-sectional nature, arbitrary definitions of time-to-event, and lack of generalizability across different types of higher education institutions. Since the hazard modeling technique does not suffer from many of these deficiencies, it may be a more appropriate reporting mechanism.

The hazard model approach also permits analysts to examine whether factors influencing student dropout or graduation vary by initial year of enrollment. This "cohort" or "panel" analysis is especially useful in determining whether there are "fixed effects" due to an external event, such as a change in admissions policy. For instance, in the fall of 1987 a policy initiative known as Commitment to Focus (CTF) was implemented at the University of Minnesota. The objective was to increase the quality of the student body by being more selective in admissions and

reducing the number of undergraduate students. By studying cohorts before and after the imposition of CTF, we may be better able to understand the effect that such a policy had on student outcomes. We are now in the process of testing whether there are differences within and among the cohorts of students entering before and after this policy change.

Studying student careers longitudinally may also help save institutional resources. Most higher education funding is done through historical average cost formulas. Since higher education is most likely characterized by economies of scale, the marginal cost is probably less than average cost over some range of output. When enrollments are increasing rapidly (as in the 1960s) an institution is overcompensated if average costs exceed incremental costs. The dividend received from overcompensation may encourage institutional administrators to promote growth in enrollments. Given the trend toward greater accessibility, it may be in the social interest for institutions to be motivated toward growth. If this methodology permits us to reduce the number of students who withdraw, institutions may be able to stabilize and even increase student enrollments.

During declining enrollments there is concern that the marginal savings associated with enrolling fewer students will be lower than the savings calculated on the basis of historical average costs. If this is true, and institutions experiencing declining enrollments lose funds based on historical average cost, then they will be put in financial distress. Again, this situation may have adverse consequences for diversity, accessibility, and student quality. If the application of hazard models can improve our understanding of student attrition, administrators may be able to reduce dropout related resource losses.

Another way hazard model research may help us save resources is by being able to determine at risk students using only institutional data. The variables used in our research were obtained from data readily available at most institutions. No time consuming and expensive surveying was necessary. We realize that we probably lose the structural detail of other studies by not having traditional measures of academic and (especially) social integration, however, there are also costs associated with collection of this attitudinal data. For example, Cabrera et al. (1993) had an initial sample of 2,453 students. After surveying to get measures of academic and social integration, their effective sample dropped to 466. Even though they found the characteristics of respondents and nonrespondents to "mirror the target population in most factors" (1993, p.579), there was probably some loss of statistical precision. For instance, they have no way of telling whether unobserved characteristics caused students to self select themselves into their study. It appears that their perceived gains in structural detail were paid for by a reduction in sample size

and statistical precision. In our study we were able to include *all* new freshman matriculating in 1985. Even though we may have lost some structural detail by not including attitudinal measures, we are able to statistically control for variables that are unobserved or unmeasured. More research needs to be done to evaluate the tradeoffs between precision and structural detail.

Section VI: Future Directions and Conclusions

Our application of hazard models to student attrition is an ongoing research project. Some areas we are currently examining are: redefining the definition of stopout; estimation of a repeated events model, specification of a competing risks model, doing policy simulations, and calculating the expected rate of attrition for various subgroups within the institution.

We will soon be able to specify models that relax the simplifying definition of stopout used for this paper. We will define the dependent variable as "leaving the institution for two consecutive terms" and will also examine how the results are affected when stopout is defined as "missing three terms in a row". Once we are comfortable with the above analyses we will specify a repeated events model that will permit us to examine whether the factors affecting multiple exits differ from those affecting the first stopout.

Another area of interest we are exploring is a competing (correlated) risks model. This technique will permit a joint determination of the transition into different end states, for instance, graduation and dropout or stopout. Since graduation and attrition may be interdependent processes (negatively correlated) it would be interesting to examine how specific covariates affect each of these competing events. Once the basic competing risks process of attrition and graduation is understood, we plan to examine the full range of ways students exit the university; through academic dismissal, voluntary exit, transfer, and graduation. Although these methods somewhat complicate the analysis, they offer a more precise way of analyzing the retention and attrition of university students.

We would also like to explore how changes in policy related variables affect attrition rates. For instance, we think it would be useful to examine how student attrition behavior changes when financial aid is increased or decreased. This simulation could provide policy-makers with information on how financial aid and/or scholarship policies affect student attrition. Another area we are interested in is estimating expected rates of attrition/graduation based on our hazard model results. Examination of expected rates would allow us to assess how well a student (or group of students) actually did relative to what we would expect based on the regressors included in the model (Astin, 1993). This information may be valuable in determining which groups

within an institution have higher than expected risk of attrition once controlling for input and environmental factors.

In conclusion, using longitudinal models to study educational outcomes has been sparse at best. As institutional researchers we need to explore how new methods can be used to explain various types of longitudinal events that take place in higher education settings. Hopefully, the use of hazard models will be one of the tools institutional researchers use in our continuing effort to help decision-makers design better policies.

References

American Council on Education, "College Going, Persistence, and Completion Patterns in Higher Education: What Do We Know." Research Briefs, Division of Policy Analysis and Research, Volume 2, Number 3, 1991.

Astin, A.W., Assessment for Excellence: The Philosophy and Practice of Assessment and Evaluation in Higher Education Phoenix, AZ: Oryx Press, 1993.

Bean, J.P., "Dropouts and Turnover: The Synthesis and Test of a Causal Model of Student Attrition". Research in Higher Education, Vol.12: 155-187, 1980.

_____ "The Synthesis of a Theoretical Model of Student Attrition". Paper presented at the American Education Research Association's annual meeting, Los Angeles, April, 1981.

_____ "Student Dependency Theory: An Explanation of Student Retention in College". Paper presented at the Annual Meeting of Association for the Study of Higher Education, Minneapolis, October, 1992.

Cabrera, A.F., Nora, A., and Castaneda, M.B., "The Role of Finances in the Persistence Process: A Structural Model", Research in Higher Education, Vol. 33, No. 5, 1992.

_____ "College Persistence: Structural Equations Modeling Test of an Integrated Model of Student Retention", Journal of Higher Education, Vol. 64, No. 2, 1993.

Cox, D. R., "Regression Models and Life-Tables (with discussion)" Journal of the Royal Statistical Society, B, 34, 1972.

DesJardins, S.L., "Using Hazard Models to Study Student Careers", Paper Presented at the 33rd Annual Forum of the Association for Institutional Researchers, Chicago, IL, May, 1993.

Dey, E.L. and Astin, A.W., "Predicting College Student Retention: Comparative National Data from the 1982 Freshman Class", Los Angeles: University of California, Graduate School of Education, Higher Education Research Institute, 1989.

Han A. and Hausman, J., "Flexible Parametric Estimation of Duration and Competing Risk Models," Journal of Applied Econometrics, 5, 1-28, 1990.

Heckman, J. J. and Singer, B., "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data," Econometrica, 52, 271-320, 1984.

Lancaster, T., "Econometric Methods for the Duration of Unemployment," Econometrica, 47, 939-956, 1979.

McCall, B., "Tesintg the Proportional Hazards Assumption in the Presence of Unmeasured Heterogeneity," Journal of Applied Econometrics, (forthcoming).

Meyer, B., "Semi-Parametric Estimation of Hazard Models," M.I.T. working paper, 1986.

_____ "Unemployment Insurance and Unemployment Spells," Econometrica, 58,757-782, 1990.

Ottinger, C., "College Going, Persistence, and Completion Patterns in Higher Education: What Do We Know?", American Council on Education, Division of Policy Analysis and Research, Research Briefs, Volume 2, Number 3, 1991.

Pascarella, E. T., "Student-Faculty Informal Contact and College Outcomes". Review of Educational Research 50: 545:595, 1980.

_____ Studying Student Attrition. New Directions for Institutional Research, San Francisco: Jossey-Bass, 1982.

_____ "A Program for Research and Policy Development on Student Persistence at the Institutional Level". Journal of College Student Personnel, 27: 100-107, 1986.

Pascarella, E. T., and Terenzini, P. T., "Predicting Freshman Persistence and Voluntary Dropout Decisions from a Theoretical Model", Journal of Higher Education, 51: 60-75, 1980.

_____ How College Affects Students: Findings and Insights from Twenty Years of Research. San Francisco, CA: Jossey-Bass, 1991.

Prentice, R. and Gloeckler, J., "Regression Analysis of Grouped Survival Data with Application to Breast Cancer Data," Biometrics, 34, 57-67, 1978.

Price, J.L., The Study of Turnover. Ames, Iowa: Iowa State University Press, 1977.

Singer, J.D. and Willett, J.B., "Modeling the Days of Our Lives: Using Survival Analysis When Designing and Analyzing Longitudinal Studies of Duration and the Timing of Events", Psychological Bulletin, Vol. 110, Number 2: 268:290, September 1991.

_____ It's About Time: Using Discrete-Time Survival Analysis To Study Duration and the Timing of Events, Unpublished Manuscript , March 1992.

Spady, W., "Dropouts from Higher Education: An Interdisciplinary Review and Synthesis", Interchange, Vol. 1: 64:85, 1970.

Stage, F.K., "University Attrition: LISREL with Logistic Regression for the Persistence Criterion", Research in Higher Education, Vol. 29, No. 4: 343:357, 1988.

Tinto, V., "Dropout from Higher Education: A Theoretical Synthesis of Recent Research", Review of Educational Research, Vol. 45: 89-125, 1975.

_____ "Stages of Student Departure: Reflections on the Longitudinal Character of Student Leaving." Journal of Higher Education, Vol.59: 438-455, 1988.

Willett, J. B. and Singer, J. D., "From Whether to When: New Methods for Studying Student Dropout and Teacher Attrition." Review of Educational Research, Vol. 61, No. 4: 407-450, Winter 1991.